R E P O R T R E S U M E S

ED 016 292 EA 001 068

METHODOLOGY FOR AN EDUCATIONAL ATTAINMENT MODEL.

BY- WISLER, CARL E.
NATIONAL CENTER FOR EDUCATIONAL STATISTICS (DHEW)

REPORT NUMBER TN-46 PUB DATE 5 SEP 67
EDRS PRICE MF-$0.25 HC-$0.80 18P.

DESCRIPTORS- *METHODOLOGY, *ACADEMIC ACHIEVEMENT, *MODELS, AGE GROUPS, STATISTICAL ANALYSIS, CHARTS, DISTRICT OF COLUMBIA,

THIS NOTE EXAMINES EDUCATIONAL ATTAINMENT AS A SOCIAL INDICATOR AND SUGGESTS A METHODOLOGY FOR DEALING WITH THE INHERENT PROBLEMS. THREE MEASURES OF EDUCATIONAL ATTAINMENT ARE CONSIDERED FOR THE MODEL--ACCUMULATED EDUCATIONAL ATTAINMENT (AEA), CURRENT EDUCATIONAL ATTAINMENT (CEA), AND AGGREGATE EDUCATIONAL CAPACITY. THE CEA MEASURE IS USED. THE MODEL RESTS ON THE ASSUMPTION THAT WHEN INDIVIDUALS ARE AGGREGATED, A FEW MAJOR FACTORS, WHICH MAY BE REGRADED FROM A NEGATIVE POINT OF VIEW AS BARRIERS TO EDUCATION, WILL EXPLAIN VARIATION IN ATTAINMENT. THESE FACTORS WILL VARY IN IMPORTANCE ACCORDING TO THE STAGES OF THE EDUCATIONAL SYSTEM, BUT THEY WILL ALL HAVE AN EVENTUAL EFFECT UPON THE CEA CURVE. CAUSAL INFERENCES AND A DYNAMIC STRUCTURE FOR EDUCATIONAL ATTAINMENT ARE ALSO DISCUSSED. (HW)

ED016292

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

Working Paper
Does Not Reflect
Official Policy of
The Office of Education

NATIONAL CENTER FOR EDUCATIONAL STATISTICS
Division of Operations Analysis

METHODOLOGY FOR AN EDUCATIONAL ATTAINMENT MODEL

by

Carl E. Wisler

Technical Note
Number 46

September 5, 1967

EA 001 068

OFFICE OF EDUCATION/U.S. DEPARTMENT OF HEALTH, EDUCATION, AND WELFARE

NATIONAL CENTER FOR EDUCATIONAL STATISTICS
Alexander M. Mood, Assistant Commissioner

DIVISION OF OPERATIONS ANALYSIS
David S. Stoller, Director

METHODOLOGY FOR AN EDUCATIONAL ATTAINMENT MODEL

The Problem

Educational attainment is widely used as a measure of the output of our educational enterprise. As such it seems likely to take a permanent position in the growing list of social indicators which may some day be comparable to economic indicators as measures of the well-being of the nation. These social indicators should not, however, be regarded only as a record of past performance and current status; they can also provide the basis of planning and action. Especially for the latter purpose it is desirable to have a clear understanding of what the indicator is actually measuring and, if possible, of the factors which cause it to behave the way it does. Without causal interpretation, social indicators will be of little use in choosing socially desirable courses of action. The purpose of this note is to explore these considerations with respect to educational attainment as a social indicator and to suggest a methodology for dealing with the problem.

The metric used almost universally for educational attainment is the years of school completed. Such information is collected by the Bureau of the Census and for the years 1940, 1950, and 1960 is illustrated in Figure 1. The curves show the per cent of the U.S. population 25 years and older which had completed at least a given level of education. For example, in 1960, 41% of the U.S. population 25 years and older had completed four years of high school or some higher level of education. The measure of educational attainment with this population base will subsequently be referred to as the accumulated educational attainment (AEA). Because many of the people in this age group have long since completed their formal education, AEA is slow to respond to possibly major changes in society's behavior with respect to education.

A more responsive indicator, which will be called the current educational attainment (CEA), is illustrated in Figure 2. The CEA base population is 24-year-olds, the oldest single age group for which data is readily available. The choice of the 24-year-old represents a compromise between a lower age which would be more responsive to changes at the elementary and secondary levels and a higher age which would be more indicative of total college attainment. A CEA series on the latter can be misleading unless it is noted that the age distribution of, for example, college graduates changes with time. Thus in 1950 the median age at college graduation was 23.6 while in 1960 it was 22.9.[1/] The confounding effect of a changing age distribution needs further study.

1/ U.S. Bureau of the Census. Current Population Reports, Series P-23, No. 9, November 8, 1963.

FIGURE 1 - Accumulated Educational Attainment

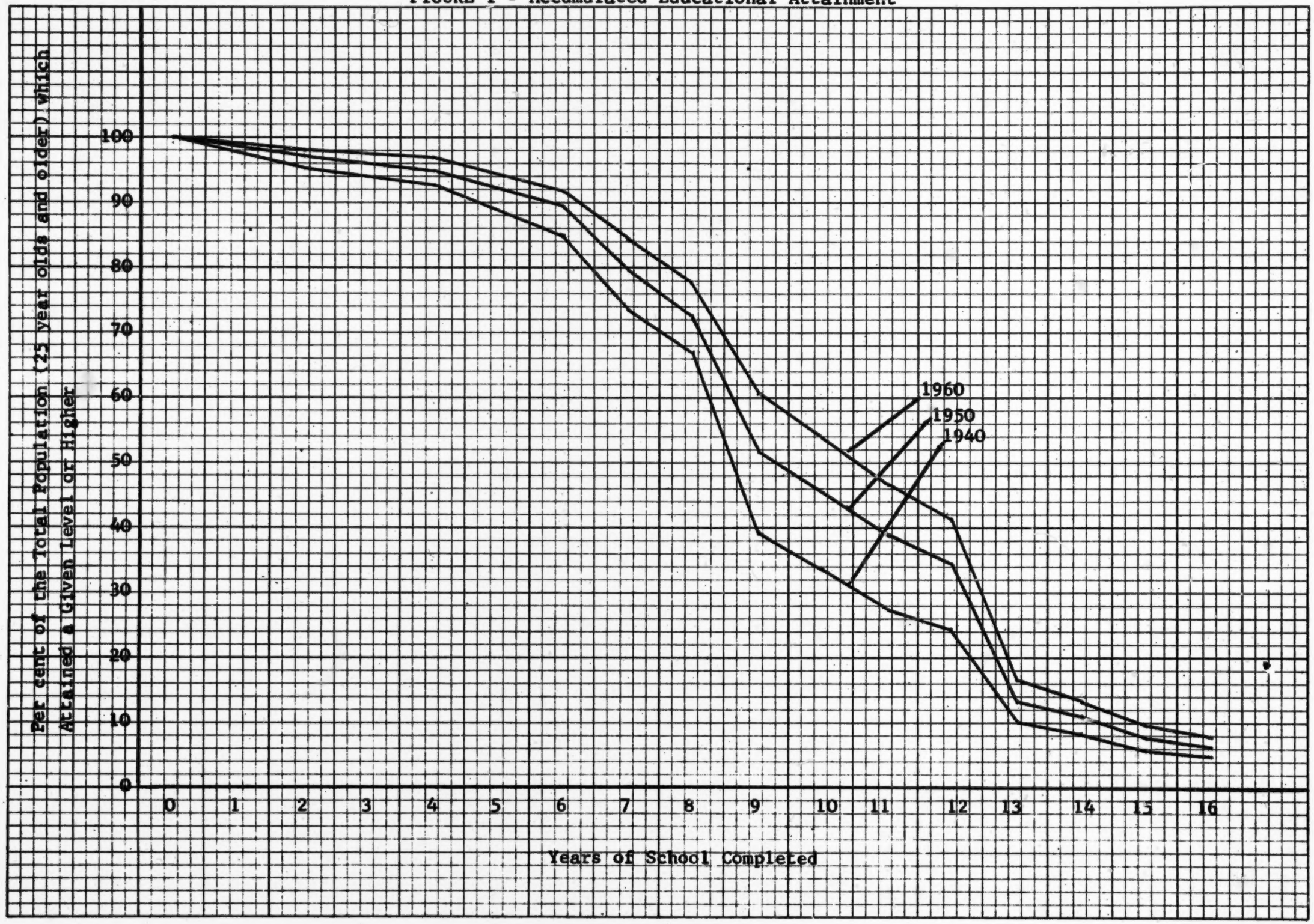
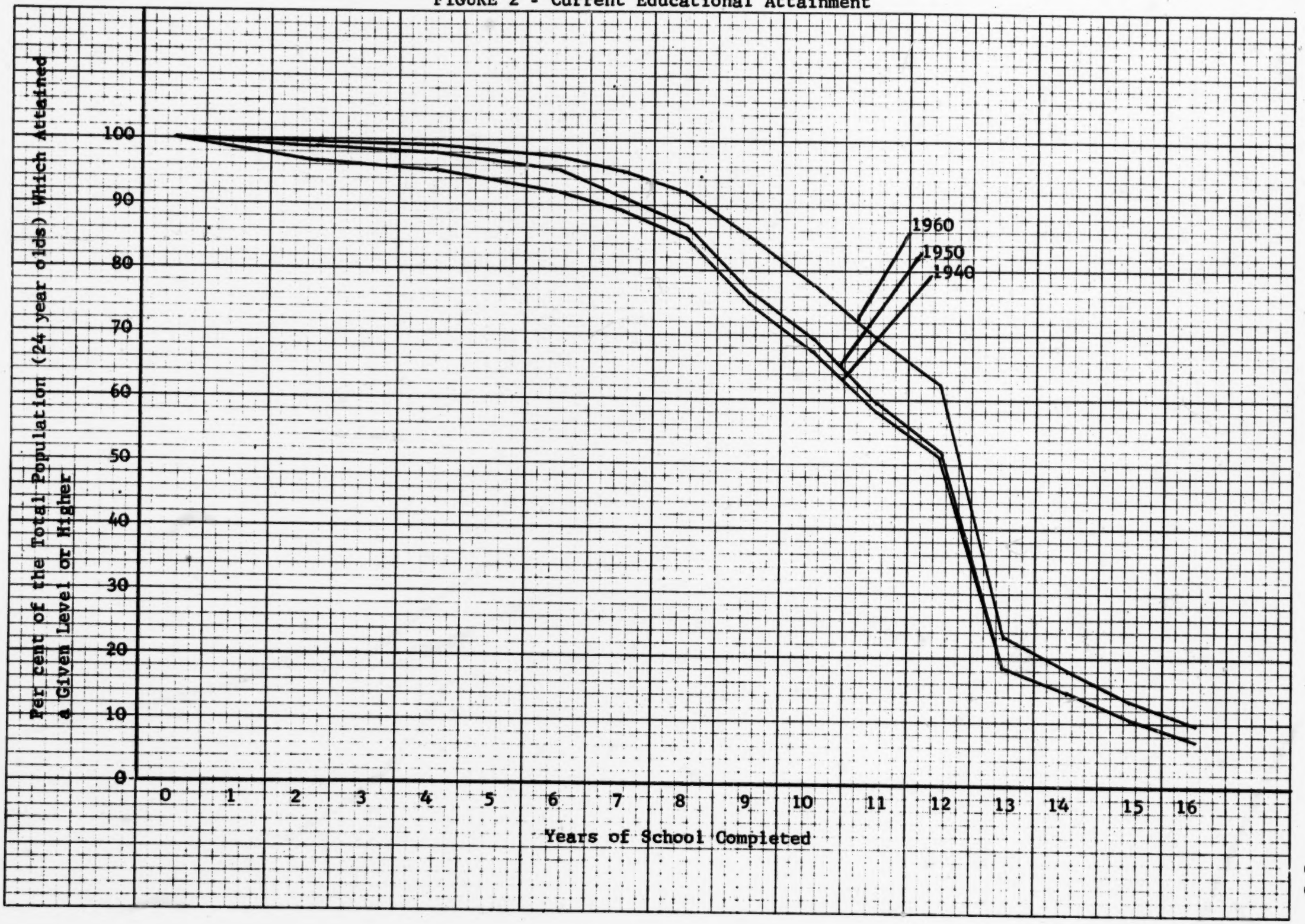


FIGURE 2 - Current Educational Attainment



A third aspect of educational attainment which will be of interest is the aggregate potential of the nation's population, that is, its educational capacity. For present purposes capacity will be defined as the years of school which could be completed if all barriers to acquiring an education other than innate intelligence were removed. Though the idea of a well-defined upper limit to any given individual's educability is probably neither a tractable nor desirable concept, an aggregate measure may be feasible and useful. Just as in the mechanics of gases, the behavior of the statistical aggregate may for many purposes be more interesting than that of individuals.

Three different measures of educational attainment have been put forth and all are in terms of the "years of school completed." Though there are several aspects of this metric which detract from its suitability, it appears to be the best for which data is readily available. Some discussion of errors in the census data and procedures for adjustment may be found in Folger and Nam[2/] and Orr and Nam[3/]. The illustrated curves in Figures 1 and 2, however, are unadjusted and a more detailed consideration of the measurability problems of educational attainment is outside the scope of this note.

Having introduced the concepts of CEA, AEA, and capacity, we would now like to see how they may be used as a basis for further developments. Since it is possible to relate CEA and AEA quantitatively only the former will be considered in detail here. A relationship between the two is given in the Appendix. Further developments of the capacity concept will be reserved for a later note.

An Attainment Model

Numerous and diverse factors affect individual progress through the formal educational system. The discussion which follows, however, rests on the assumption that when individuals are aggregated a few major factors, which may be regarded from a negative point of view as barriers to education, will explain variation in attainment. These factors will vary in importance according to the stage of the educational system but they will all have an eventual effect upon the CEA curve. This observation suggests that a suitable model might involve a set of relationships between attainment and the various barrier factors. For example, postulating a linear relationship between the

---

2/ J. K. Folger and C. B. Nam, "Educational Trends from Census Data," Demography, Vol. 1, No. 1, 1964, pp. 247-257.

3/ S. B. Orr and C. B. Nam, "Estimates of the 'True' Educational Distribution of the Adult Population of the United States from 1910 to 1960," Paper prepared for the annual meeting of the Population Association of America, Cincinnati, Ohio, April 28-29, 1967.

i^{th} attainment level[4/] and the explanatory variables, we could have the following type of structure

$$\sum_{j=1}^{n} \beta_{ij} x_j + x_i' = e_i \quad i = n + 1 \tag{1}$$

$$\sum_{j=1}^{n} \beta_{ij} x_j + \beta_{i,i-1} x_{i-1} + x_i = e_i \quad i = n+2, n+3, \ldots, m$$

where x_i = proportion of the population attaining the i^{th} level (or higher)

x_j = value of the j^{th} barrier variable

e_i = unexplained residual in the i^{th} equation

β's = coefficients of the explanatory variables; to be estimated by statistical procedures.

In this formulation, the x_j variables, n in number, are regarded as exogenous where an exogenous variable is one whose value is statistically independent of the values of all random disturbances in the model. [5/] Variables which are not exogenous are endogenous and it is those variables for which we wish to develop causal explanations.

4/ In the mathematical formulation there will be no attempt to make the numerical indicator of level correspond to any conventions in the U.S. educational system.

5/ C. Christ, Econometric Models and Methods, (New York: John Wiley, 1966, Pp. 156-157).

Equation (1) can be written in matrix form as

(2) $\beta_{ex} X_{ex} + \beta_{en} X_{en} = e$

where X_{ex} = an n dimensional column vector of exogenous (barrier) variables

X_{en} = an m dimensional column vector of endogenous variables

β_{ex} = an m by n matrix of coefficients of exogenous variables

β_{en} = an m by m matrix of coefficients of endogenous variables

e = an m dimensional column vector of residuals

It may be noted that the matrix β_{en} is triangular with ones along the principal diagonal. Before examining the properties of this mathematical construct we supply a particular model.

Consider as endogenous variables just three salient levels on the attainment curve, viz. high school graduates, two years of college completed and four years of college completed.

Let x_1 = a measure of motivation for attaining education

x_2 = proximity to an accredited two- or four-year college

x_3 = ability to pay for a college education

x_4 = proximity to an accredited four-year college

x_5 = proportion of the population which has completed the eighth grade

x_6 = proportion of the population which has completed high school

x_7 = proportion of the population which has completed two years of college

x_8 = proportion of the population which has completed four years of college.

A possible explanatory model for the endogenous variables x_6, x_7, x_8 in terms of the exogenous variables x_1, x_2, x_3, x_4, x_5 is 6/.

$$\text{(3a)}\quad \beta_{61}x_1 + \beta_{65}x_5 + x_6 = e_6$$

$$\text{(3b)}\quad \beta_{71}x_1 + \beta_{72}x_2 + \beta_{73}x_3 + \beta_{76}x_6 + x_7 = e_7$$

$$\text{(3c)}\quad \beta_{81}x_1 + \beta_{83}x_3 + \beta_{84}x_4 + \beta_{87}x_7 + x_8 = e_8$$

The interpretation of equation (3b), for example, is that the proportion of the population completing two years of college depends, by postulate, upon four explicit factors: the proportion completing high school, a motivational factor, the proximity to a college and the ability to pay plus some unspecified factors the effect of which is summarized in the error term e_7. More generally, in each equation the variable with a coefficient of unity is regarded as a dependent variable which is "explained" by the dependent variable of the immediately preceding equation (when there is one) and some combination of the barrier variables. Note that if x_8 is measured in the year t, x_7 is associated with the year t-2, x_6 with year t-4, and x_5 with year t-8.

The foregoing system of relations is a set of simultaneous equations the parameters of which we wish to estimate. In general, fitting systems of simultaneous equations to observed data introduce statistical problems which do not arise when only single equation relationships are considered. Fortunately, the system of interest here is a special type known as a recursive system and it is possible to avoid the problems associated with more general structures. A simultaneous equation model is said to be recursive if the β_{en} matrix is triangular and the covariance matrix of the error terms is diagonal. Under these assumptions the parameters of a recursive system may be estimated by the method of ordinary least squares applied to each equation in turn. The estimators so obtained are consistent, unbiased and, if the errors are normally distributed, equivalent to maximum likelihood estimators. 7/

6/ It will be assumed that all variables are expressed as deviations from their respective means.

7/ E. Malinvaud, Statistical Methods of Econometrics, (Rand McNally, 1966).

Causal Inference

The special form of the model thus lends itself to a simple statistical analysis, but the resulting parameter estimates will have a causal interpretation only if we are willing to accept (3a-c) as an appropriate causal model. By causal interpretation we mean that the following type of statement is valid: If, in the real world, the x_2 variable undergoes a change of one unit this will produce a change of β_{72} units in x_7, assuming that all other variables in equation (3b) are held constant. The potential usefulness of this kind of statement to policymakers is evident. Once the parameters are known, if the desired social goal is a certain change in attainment, x_7, the model would permit experimentation with the barrier variables in an effort to effect the desired change.[8/] In addition, by using subsequent equations in the recursive system, the propagated effects of the proposed changes could be estimated. If, on the other hand, the relationship between x_2 and x_7 reflects only correlation without causality, the model is of much less use to a policymaker though it may be valuable to a researcher.

The potentiality of the recursive model of educational attainment as an aid to policy making thus depends upon the possibilities of making inferences about alternative causal hypotheses. One such hypothesis might be represented by the system (3a-c). A causal diagram of the system facilitates interpretation of the hypothesis and is given in Figure 3. In the diagram an arrow from x_i to x_j, for example, indicates a causal link from x_i to x_j and correspondingly a non-zero value for β_{ji} in (3a-c). Conversely, the absence of an arrow would imply that $\beta_{ji} = 0$. It turns out that additional information of this type plus the assumption that the covariance matrix of the error terms is diagonal may be used to judge the appropriateness of alternative hypotheses.

8/ A logical extension of the experimental approach would be to employ a mathematical programming model in which the objective function would be, say, an expression of the costs associated with changes in the barrier variables and an inequation of the type $\sum_{j=1}^{i-1} -\beta_{ij}x_j \geq L_i$ would be one of the constraints.

L_i is the attainment desired at level i.

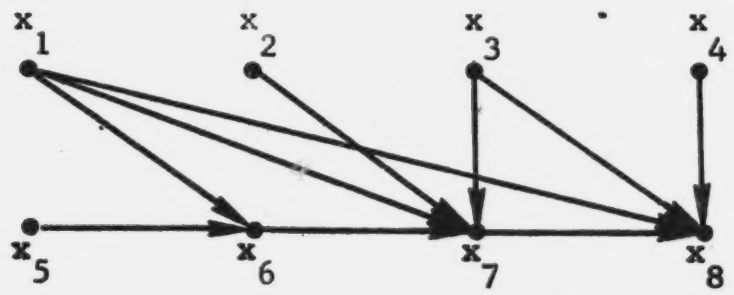


FIGURE 3

To see how this may be done we proceed by using an approach suggested by Boudon. 9/ First, consider the foregoing model in matrix form.

(4) $$\beta X = e$$

where

$$\beta = \begin{bmatrix} \beta_{61} & 0 & 0 & 0 & \beta_{65} & 1 & 0 & 0 \\ \beta_{71} & \beta_{72} & \beta_{73} & 0 & 0 & \beta_{76} & 1 & 0 \\ \beta_{81} & 0 & \beta_{83} & \beta_{84} & 0 & 0 & \beta_{87} & 1 \end{bmatrix}$$

$$X' = (x_1 \quad x_2 \quad x_3 \quad x_4 \quad x_5 \quad x_6 \quad x_7 \quad x_8)$$

$$e' = (e_6 \quad e_7 \quad e_8)$$

Postmultiplying (4) by X' and taking expectations we obtain

$$E(\beta XX') = E(e\,X')$$

or

(5) $$\beta E(XX') = E(e\,X')$$

9/ R. Boudon, "A Method of Linear Causal Analysis: Dependence Analysis," American Sociological Review, Vol. 30, No. 3, June 1965, pp. 365-374.

Note that $E(XX')$ can be expressed in terms of standard deviations and correlation coefficients since $E(X_iX_j) = \rho_{ij}\ \sigma_i\ \sigma_j$. Equation (5) has the two types of a priori restrictions alluded to earlier. First, the presence of zeros in the β matrix represent the absence of causal links between certain variables. In other words the pattern of zeros corresponds to one postulated causal explanation of the variables x_6, x_7, and x_8. Our aim is to test the appropriateness of this particular hypothesis. The second a priori constraint on (5) results from the diagonal form of the error covariance matrix. The definition of exogenous variable and the diagonal form of the covariance matrix imply that the covariance of an error term with a variable which has a subscript smaller than that of the error term is zero. The covariance matrix is then of the form

$$E(eX') = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & * & * & * \\ 0 & 0 & 0 & 0 & 0 & 0 & * & * \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & * \end{bmatrix}$$

where * indicates a value not identically zero. The equations of (5) which will be important are those corresponding to the 18 zeros in the above matrix. The first of these is

$$\beta_{61}\ \sigma_1^2 + \beta_{65}\ \sigma_1\ \sigma_5\ \rho_{15} + \sigma_1\ \sigma_6\ \rho_{16} = 0$$

which may be rewritten as

$$\left(\beta_{61}\ \frac{\sigma_1}{\sigma_6}\right)\ \sigma_1\ \sigma_6 + \left(\beta_{65}\ \frac{\sigma_5}{\sigma_6}\right)\ \sigma_1\ \sigma_6\ \rho_{15} + \sigma_1\ \sigma_6\ \rho_{16} = 0$$

and letting

$$\beta'_{ij} = \beta_{ij}\ \frac{\sigma_j}{\sigma_i}$$

$$\left(\beta'_{61} + \beta'_{65}\ \rho_{15} + \rho_{16}\right)\ \sigma_1\ \sigma_6 = 0$$

which implies that

$$\beta'_{61} + \beta'_{65}\ \rho_{15} + \rho_{16} = 0$$

In a similar way 17 other equations can be written out giving

(6a) $\beta'_{61} + \beta'_{65}\rho_{15} + \rho_{16} = 0$

(6b) $\beta'_{61}\rho_{21} + \beta'_{65}\rho_{25} + \rho_{26} = 0$

(6c) $\beta'_{61}\rho_{31} + \beta'_{65}\rho_{35} + \rho_{36} = 0$

(6d) $\beta'_{61}\rho_{41} + \beta'_{65}\rho_{45} + \rho_{46} = 0$

(6e) $\beta'_{61}\rho_{51} + \beta'_{65} + \rho_{56} = 0$

(6f) $\beta'_{71} + \beta'_{72}\rho_{12} + \beta'_{73}\rho_{13} + \beta'_{76}\rho_{16} + \rho_{17} = 0$

(6g) $\beta'_{71}\rho_{21} + \beta'_{72} + \beta'_{73}\rho_{23} + \beta'_{76}\rho_{26} + \rho_{27} = 0$

(6h) $\beta'_{71}\rho_{31} + \beta'_{72}\rho_{32} + \beta'_{73} + \beta'_{76}\rho_{36} + \rho_{37} = 0$

(6i) $\beta'_{71}\rho_{41} + \beta'_{72}\rho_{42} + \beta'_{73}\rho_{43} + \beta'_{76}\rho_{46} + \rho_{47} = 0$

(6j) $\beta'_{71}\rho_{51} + \beta'_{72}\rho_{52} + \beta'_{73}\rho_{53} + \beta'_{76}\rho_{56} + \rho_{57} = 0$

(6k) $\beta'_{71}\rho_{61} + \beta'_{72}\rho_{62} + \beta'_{73}\rho_{63} + \beta'_{76} + \rho_{67} = 0$

(6l) $\beta'_{81} + \beta'_{83}\rho_{13} + \beta'_{84}\rho_{14} + \beta'_{87}\rho_{17} + \rho_{19} = 0$

(6m) $\beta'_{81}\rho_{21} + \beta'_{83}\rho_{23} + \beta'_{84}\rho_{24} + \beta'_{87}\rho_{27} + \rho_{29} = 0$

(6n) $\beta'_{81}\rho_{31} + \beta'_{83} + \beta'_{84}\rho_{34} + \beta'_{87}\rho_{37} + \rho_{39} = 0$

(6o) $\beta'_{81}\rho_{41} + \beta'_{83}\rho_{43} + \beta'_{84} + \beta'_{87}\rho_{47} + \rho_{49} = 0$

(6p) $\beta'_{81}\rho_{51} + \beta'_{83}\rho_{53} + \beta'_{84}\rho_{54} + \beta'_{87}\rho_{57} + \rho_{59} = 0$

(6q) $\beta'_{81}\rho_{61} + \beta'_{83}\rho_{63} + \beta'_{84}\rho_{64} + \beta'_{87}\rho_{67} + \rho_{69} = 0$

(6r) $\beta'_{81}\rho_{71} + \beta'_{83}\rho_{73} + \beta'_{84}\rho_{74} + \beta'_{87} + \rho_{79} = 0$

The foregoing 18 equations can be used to test the consistency of the proposed model with the observed data on the variables. [10/] This is done by computing the left side of (6a-r) (from the results of a least squares regression applied equation by equation to (4)) and comparing the value thus obtained to zero. Though intuitively we know that the larger the deviation the worse the fit of the model to the data, statistical tests of goodness of fit should be developed.

At this point some references to other work in linear causal analysis are in order. Coefficients of the type - β'_{ij} were called path coefficients by Wright in his early work on causal analysis. [11,12/] They were apparently regarded as a means of interpreting causal models without the formalism for testing the adequacy of the model. With respect to interpretation the path coefficient indicates the degree to which variation of the dependent variable is determined by each particular causal variable.

Following a method of analysis suggested by Simon [13/] and Blalock [14/], Boudon[15/] points out that estimates of the β'_{ij} coefficients can be obtained by solving the system of equations (6). For example, given values for the correlation coefficients, (6a) and (6b) can be solved simultaneously for β'_{61} and β'_{65}; (6f-i) for β'_{71}, β'_{72}, β'_{73}, and β'_{76}; and so on. The values obtained in this way are called standardized dependence coefficients by Boudon. Solving (6) for dependence coefficients would "use up" ten equations leaving eight with which to test the model. There is, however, no a priori rule for deciding which ten equations to use in solving for the unknowns and consequently the dependence coefficients do not have unique values. The ambiguity of the method seems to make it less preferable than applying least squares to (4).

10/ In general if there are r endogenous variables and s exogenous variables there will be $rs + \frac{r(r-1)}{2}$ such equations.

11/ S. Wright, "On the Nature of Size Factors," Genetics, Vol. 3, 1918, pp. 367-374.

12/ S. Wright, "The Method of Path Coefficients," Annals of Mathematical Statistics, Vol. V, 1934, pp. 161-215.

13/ H. Simon, "Spurious Correlation: A Causal Interpretation," Journal of the American Statistical Association, Vol. 49, Sept. 1954.

14/ H. Blalock, Jr., Causal Inferences in Nonexperimental Research, (Chapel Hill: The University of North Carolina Press, 1961).

15/ R. Boudon, op.cit.

Appendix

A Dynamic Structure for Educational Attainment

The change in the distribution of educational attainment over time may be considered in terms of a network flow model. The matrix equations corresponding to a simple flow diagram will be developed and then a general form will be stated. The relationship between CEA and AEA may be inferred from the general matrix equation.

In the network depicted in Figure 1, the nodes represent population groups classified with respect to age and level of education; the arrows indicate annual transitions between nodes.

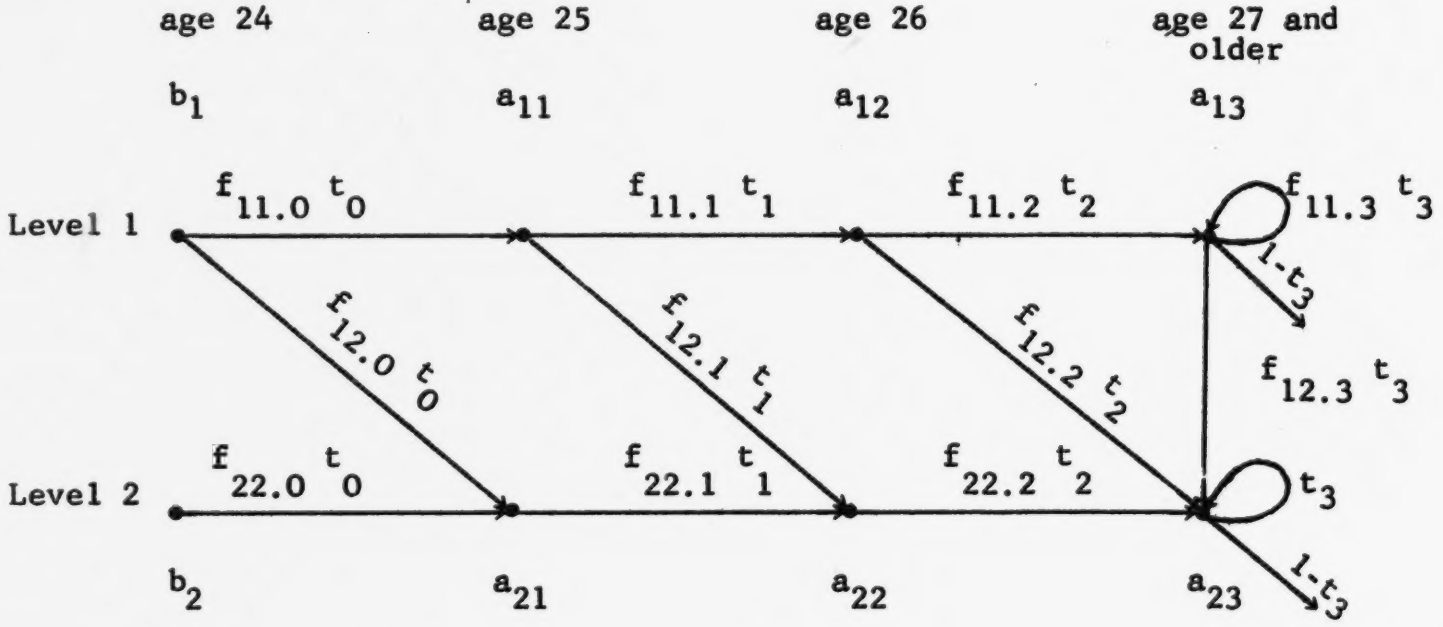


Figure 1

The following notation applies

b_i = number of 24 year olds at attainment level i.

a_{ij}= number of individuals at the i^{th} attainment level in the j^{th} age group; j=1 corresponds to the 25 year olds and so on.

$f_{ik.j}$ = proportion of individuals in the j^{th} age group who begin the year at attainment level i and end the year at level k; $f_{ik.j}$ is identically zero for k not equal to i or i + 1.

t_j = survival ratio for individuals in the j^{th} age group; j = 0 corresponds to 24 year olds, j = 1 corresponds to 25 year olds and so on.

The distribution of attainment for the population 25 years and older with respect to age and level of education is then described by the values of the a_{ij}. Equations may now be written expressing the distribution of attainment in terms of the distribution of the preceding year. When it is necessary to distinguish between the old distribution and the new one, the latter will be denoted by primes, i.e., a'_{ij}. Corresponding to Figure 1 we would then have the following result:

$$\begin{bmatrix} a'_{11} & a'_{12} & a'_{13} \\ a'_{21} & a'_{22} & a'_{23} \end{bmatrix} = \begin{bmatrix} f_{11.0} & 0 \\ f_{12.0} & f_{22.0} \end{bmatrix} \begin{bmatrix} t_0 & 0 \\ 0 & t_0 \end{bmatrix} \begin{bmatrix} b_1 & 0 & 0 \\ b_2 & 0 & 0 \end{bmatrix}$$

$$+ \begin{bmatrix} f_{11.1} & 0 \\ f_{12.1} & f_{22.1} \end{bmatrix} \begin{bmatrix} t_1 & 0 \\ 0 & t_1 \end{bmatrix} \begin{bmatrix} 0 & a_{11} & 0 \\ 0 & a_{21} & 0 \end{bmatrix}$$

$$+ \begin{bmatrix} f_{11.2} & 0 \\ f_{12.2} & f_{22.2} \end{bmatrix} \begin{bmatrix} t_2 & 0 \\ 0 & t_2 \end{bmatrix} \begin{bmatrix} 0 & 0 & a_{12} \\ 0 & 0 & a_{22} \end{bmatrix}$$

$$+ \begin{bmatrix} f_{11.3} & 0 \\ f_{12.3} & 1 \end{bmatrix} \begin{bmatrix} t_3 & 0 \\ 0 & t_3 \end{bmatrix} \begin{bmatrix} 0 & 0 & a_{13} \\ 0 & 0 & a_{23} \end{bmatrix}$$

$$= \begin{bmatrix} b_1 f_{11.0} t_0 & a_{11} f_{11.1} t_1 & a_{12} f_{11.2} t_2 + a_{13} f_{11.3} t_3 \\ b_1 f_{12.0} t_0 + b_2 f_{22.0} t_0 & a_{21} f_{22.1} t_1 + a_{11} f_{12.1} t_1 & a_{12} f_{12.2} t_2 + a_{22} f_{22.2} t_2 + a_{13} f_{12.3} t_3 + a_{23} t_3 \end{bmatrix}$$

The final matrix may be verified by comparison with Figure 1.

We now adopt the following matrix notation.

$F_j = \left[f_{ik.j}\right]$ $\quad i = 1,\text{---}, m\ ;\ j = 0, 1,\text{---}, n\ ;\ k = 1, \text{---}, m$

F_j is a triangular matrix with zeros above the principal diagonal. The transpose of F_j is a stochastic matrix.

$T_j = t_j I$ $\quad j = 0, 1,\text{---}, n$

T_j is a diagonal matrix, I is the identity matrix.

$A' = \left[a'_{ij}\right]$ $\quad i = 1,\text{---}, m\ ;\ j = 1,\text{---}, n$

$$A_j = \begin{bmatrix} 0 \text{ --- } 0 & a_{1j} & 0 \text{ --- } 0 \\ \vdots \quad \vdots & \vdots & \vdots \quad \vdots \\ 0 \text{ --- } 0 & a_{mj} & 0 \text{ --- } 0 \end{bmatrix}$$

$j = 1,\text{---}, m$

A_j is an m by n matrix with only column $j + 1$ being non-zero.

$$B = \begin{bmatrix} b_1 & 0 \text{ ---- } 0 \\ \vdots & \vdots \quad \vdots \\ b_m & 0 \text{ ---- } 0 \end{bmatrix}$$

B is an m by n matrix.

The general matrix equation for the educational attainment distribution is then

$$A' = F'_0 \, T_0 \, B + \sum_{j=1}^{m} F'_j \, T_j \, A_j$$

CEA and AEA may now be expressed as vectors in terms of B and A (or A') respectively. Since CEA is given as a proportion and B is in absolute numbers we have,

$$[CEA] = \frac{1}{I_r\, BI_c}\, BI_c$$

where I_r = a unit row vector

I_c = a unit column vector.

$[CEA]$ = a row vector representation of CEA; the vector has one element for each level in the educational system.

Similarly,

$$[AEA] = \frac{1}{I_r\, AI_c}\, AI_c$$

where

$[AEA]$ = a row vector representation of AEA.